Explore Data

Aleksei Sorokin, asorokin@hawk.iit.edu, A20394300

Load data and do some final filtering

```
df <- read.csv('data/all_coaches.csv')
data_dict <- list(
              N='Name',
              GR='Games Relative',
              WP='Win Loss Percentage',
              PGR='Playoff Games Relative',
              PWP='Playoff Win Percentage',
              CC='Conference Championships',
              C='Championships',
              HOF='Hall of Fame',
              S='Sport')
names(df) <- names(data_dict)
# export stats
head(df)
```

```
##              N       GR        WP       PGR   PWP CC C HOF        S
## 1     AJ Hinch 6.308642 0.5580000 4.5454545 0.560  2 1   0 baseball
## 2  Aaron Boone 2.000000 0.6270000 1.2727273 0.500  0 0   0 baseball
## 3 Aaron Kromer 0.375000 0.3330000 0.0000000 0.000  0 0   0 football
## 4   Abe Gibron 2.625000 0.2740000 0.0000000 0.000  0 0   0 football
## 5    Adam Gase 4.000000 0.4690000 0.3333333 0.000  0 0   0 football
## 6   Adam Oates 1.585366 0.5752212 0.4375000 0.429  0 0   0   hockey
```

```
nrow(df) # total coaches
```

```
## [1] 1920
```

```
sum(df$HOF==1) # total hall of fame coaches
```

```
## [1] 257
```

```
sum(df$HOF==0) # total non hall of fame coaches
```

```
## [1] 1663
```

```
summary(df)
```

```
##              N              GR                 WP              PGR
##  Bill Russell :   2   Min.   : 0.00617   Min.   :0.0000   Min.   : 0.0000
##  Bill Stewart :   2   1st Qu.: 0.62500   1st Qu.:0.3480   1st Qu.: 0.0000
##  George Gibson:   2   Median : 1.97546   Median :0.4570   Median : 0.0000
##  Hugo Bezdek  :   2   Mean   : 3.97329   Mean   :0.4305   Mean   : 0.9025
##  Jim Mora     :   2   3rd Qu.: 5.13610   3rd Qu.:0.5299   3rd Qu.: 0.7273
##  John Russell :   2   Max.   :47.87037   Max.   :1.0000   Max.   :22.0625
##  (Other)      :1908
```

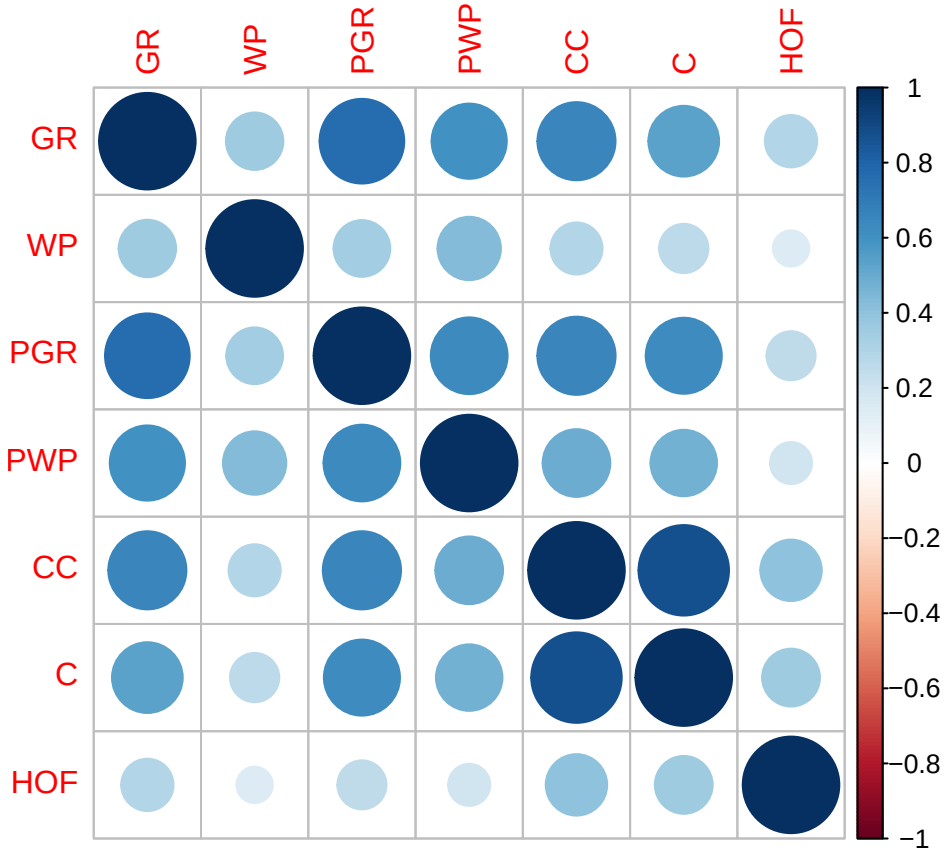
```
##       PWP                CC                 C                HOF
##  Min.   :0.0000   Min.   : 0.0000   Min.   : 0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.:0.0000
##  Median :0.0000   Median : 0.0000   Median : 0.0000   Median :0.0000
##  Mean   :0.1553   Mean   : 0.3474   Mean   : 0.2104   Mean   :0.1339
##  3rd Qu.:0.3685   3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :13.0000   Max.   :11.0000   Max.   :1.0000
##
##           S
##  baseball  :711
##  basketball:332
##  football  :500
##  hockey    :377
##
##
##
```

Correlation between data

```
library(corrplot)
```

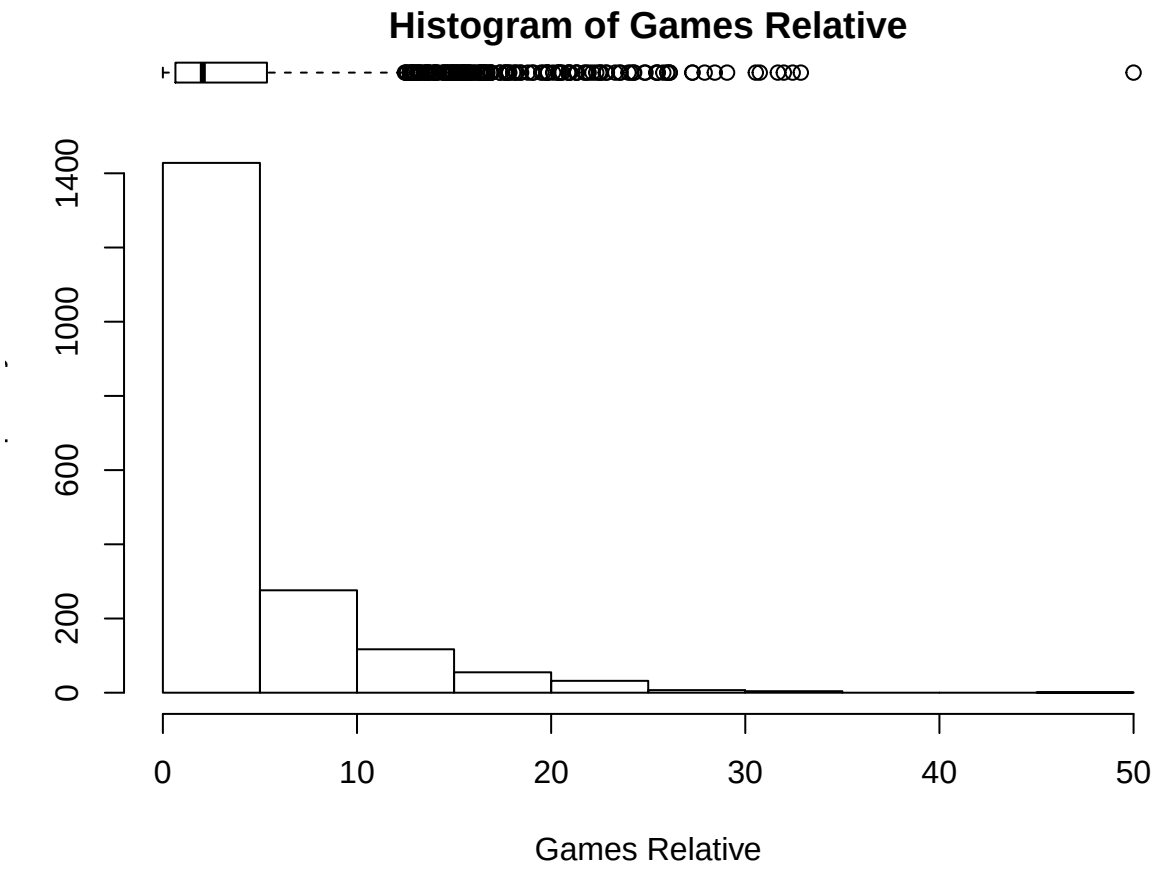
```
## corrplot 0.84 loaded
```

```
df_num <- Filter(is.numeric,df)
df_corr <- cor(df_num)
corrplot(df_corr)
```



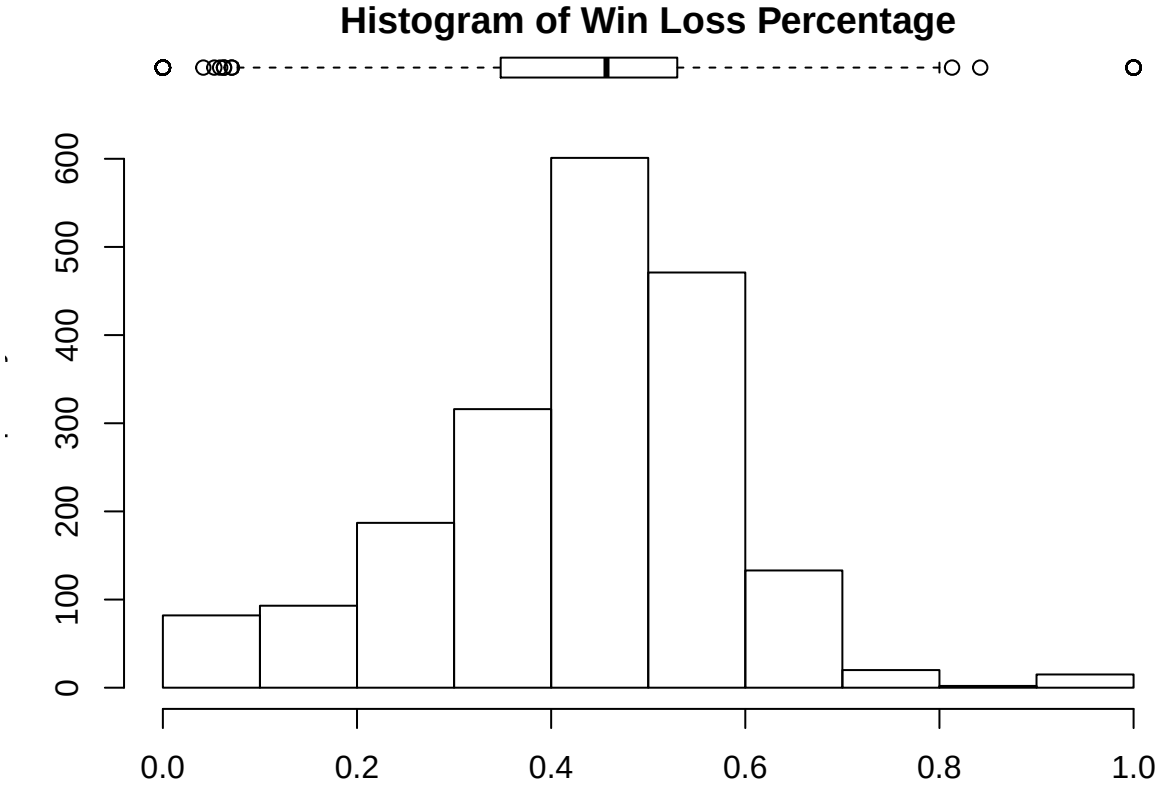
Boxplots over Histograms

```
fixed_cols <- names(df_num)[names(df_num)!='HOF']
for (col in fixed_cols){
  # gather data
  c_name <- data_dict[[col]]
  data <- df_num[[col]]
  # make plot
  layout(mat = matrix(c(1,2),2,1, byrow=TRUE),  height = c(1,8))
  par(mar=c(0, 3.1, 1.1, 2.1))
  boxplot(data , horizontal=TRUE , xaxt="n", frame=F,main=sprintf('Histogram of %s',c_name))
  par(mar=c(4, 3.1, 1.1, 2.1))
  hist(data,xlab=c_name,main='')
  # print top coaches in this category
  cat(sprintf('Top 10 coaches by %s\n',c_name))
  df_top <- df[order(df[[col]],decreasing=T),]
  print(df_top[1:10,])}
```

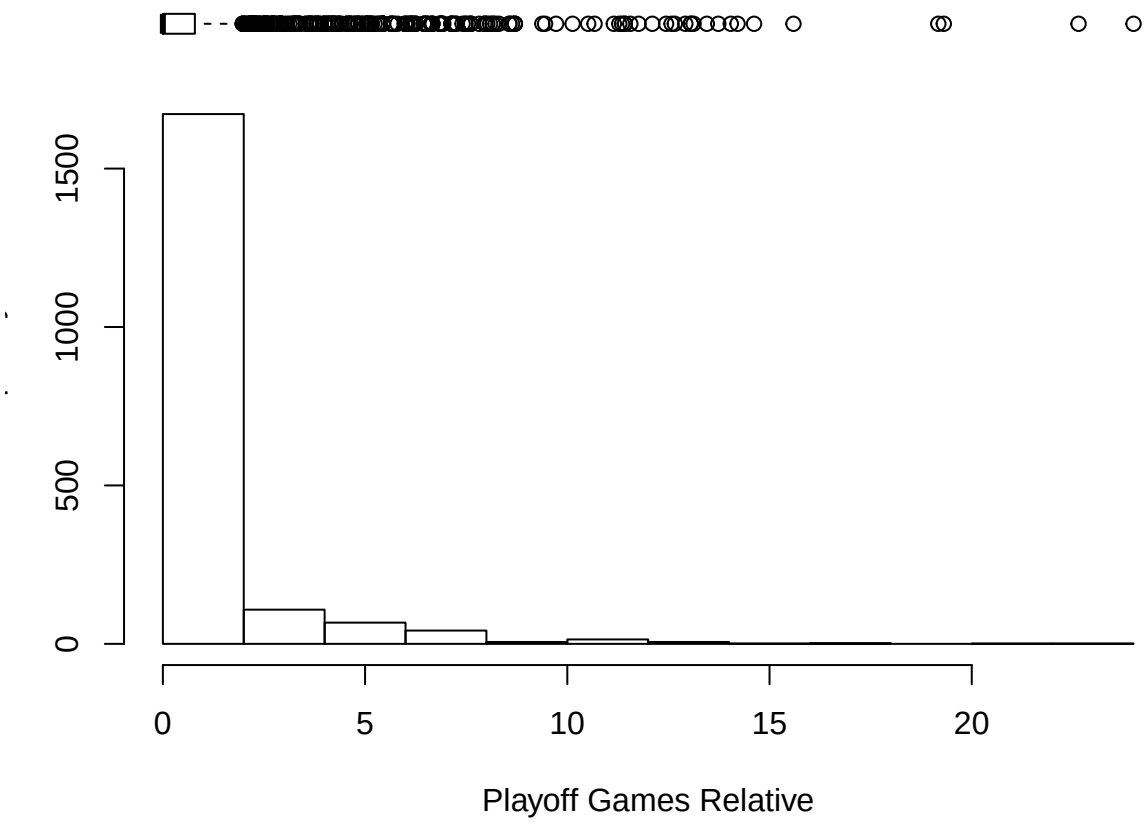



```
## Top 10 coaches by Games Relative
##                   N       GR    WP       PGR   PWP CC C HOF          S
## 403    Connie Mack 47.87037 0.486  3.909091 0.558  9 5   1   baseball
## 1856 Tony La Russa 31.46296 0.536 11.636364 0.547  6 3   1   baseball
## 754   George Halas 31.06250 0.682  3.000000 0.667  6 6   1   football
## 562      Don Shula 30.62500 0.677 12.000000 0.528  6 2   1   football
## 1277 Lenny Wilkens 30.32927 0.536 11.125000 0.449  2 1   1 basketball
## 1145   John McGraw 29.43827 0.586  4.909091 0.481 10 3   1   baseball
```

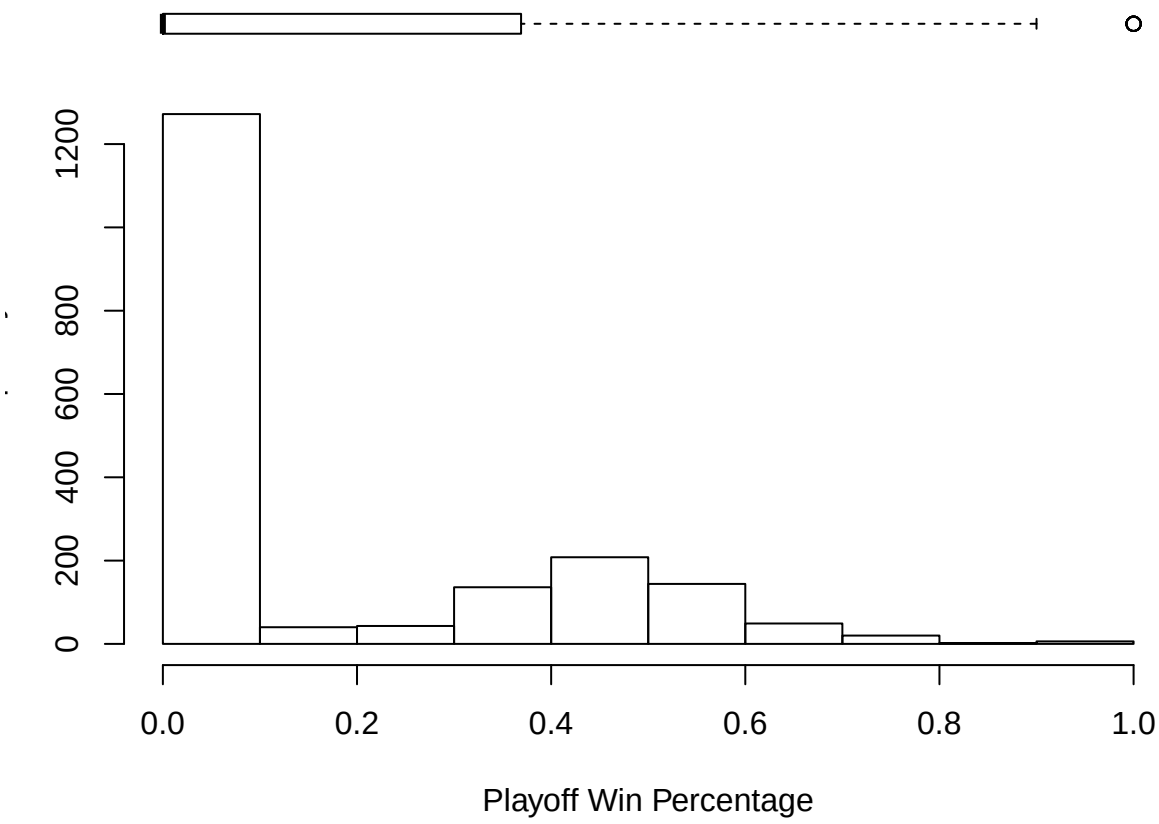
```
## 560     Don Nelson 29.24390 0.557 10.375000 0.452  0 0   1 basketball
## 236      Bobby Cox 27.82716 0.556 12.363636 0.493  5 1   1   baseball
## 287   Bucky Harris 27.22222 0.493  1.909091 0.524  3 2   1   baseball
## 1102     Joe Torre 26.72222 0.538 12.909091 0.592  6 4   1   baseball
```



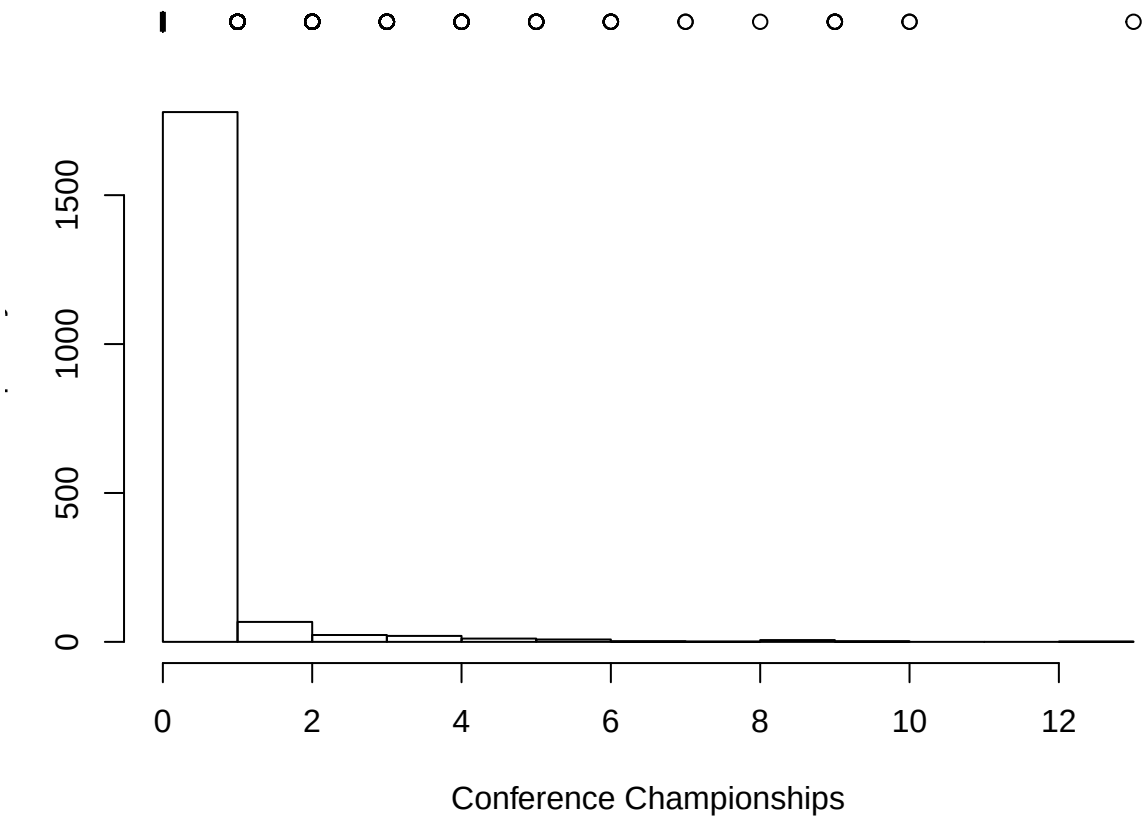

```
## Top 10 coaches by Win Loss Percentage
##                   N         GR WP PGR PWP CC C HOF        S
## 57       Andy Cohen 0.00617284  1   0   0  0 0   0 baseball
## 95        Bibb Falk 0.00617284  1   0   0  0 0   0 baseball
## 108    Bill Burwell 0.00617284  1   0   0  0 0   0 baseball
## 322      Cap Raeder 0.01219512  1   0   0  0 0   0   hockey
## 378    Chuck Drulis 0.12500000  1   0   0  0 0   0 football
## 400 Clyde Sukeforth 0.01234568  1   0   0  0 0   0 baseball
## 496      Del Wilber 0.00617284  1   0   0  0 0   0 baseball
## 526  Dick Tracewski 0.01234568  1   0   0  0 0   0 baseball
## 696     Fred Bruney 0.06250000  1   0   0  0 0   0 football
## 803      Gus Tebell 0.18750000  1   0   0  0 0   0 football
```



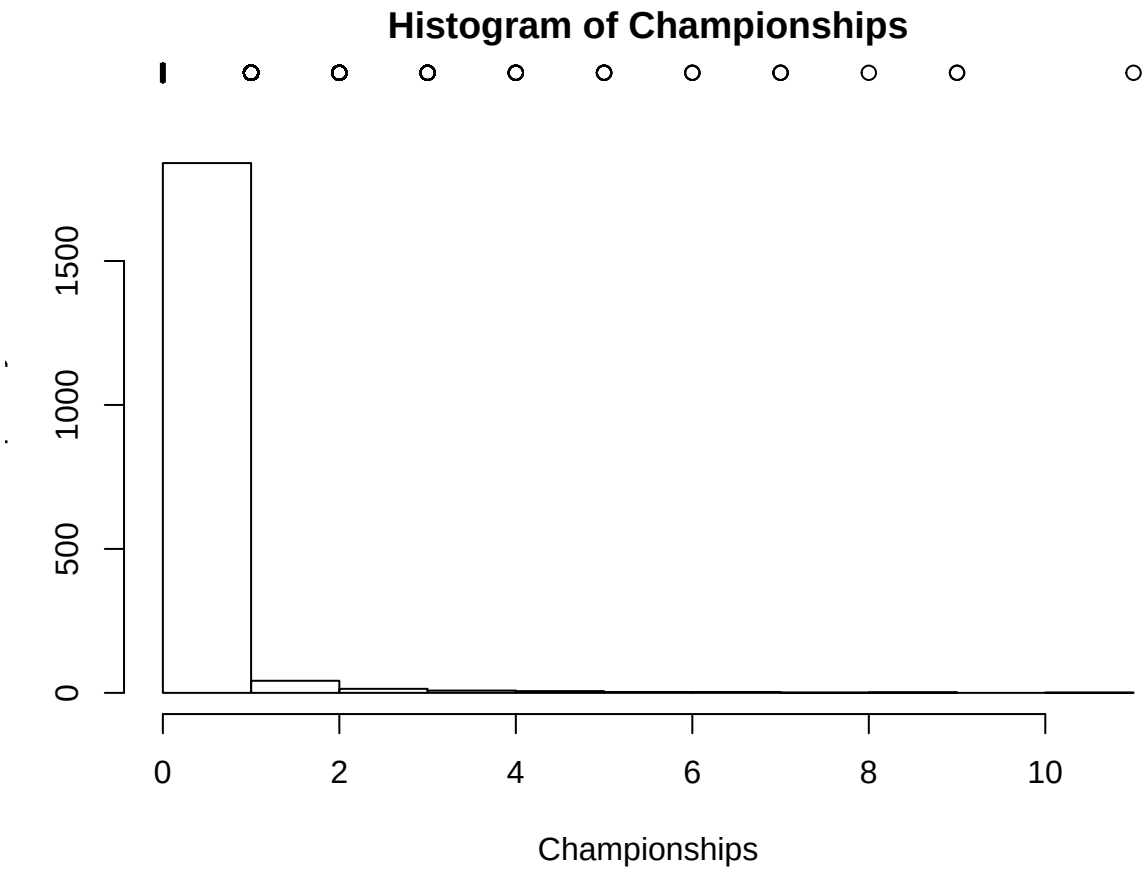

```
## Top 10 coaches by Playoff Games Relative
##                     N       GR        WP      PGR   PWP CC  C HOF          S
## 1722    Scotty Bowman 26.10976 0.6846450 22.06250 0.632  9  9   1     hockey
## 1559     Phil Jackson 20.00000 0.7040000 20.81250 0.688 13 11   1 basketball
## 795    Gregg Popovich 22.96341 0.6760000 17.75000 0.599  6  5   0 basketball
## 1514        Pat Riley 23.21951 0.6360000 17.62500 0.606  9  5   1 basketball
## 102    Bill Belichick 25.00000 0.6830000 14.33333 0.721  9  6   0   football
## 1105 Joel Quenneville 20.79268 0.6237357 13.43750 0.549  3  3   0     hockey
## 8           Al Arbour 19.59756 0.5754231 13.06250 0.589  4  4   1     hockey
## 1102        Joe Torre 26.72222 0.5380000 12.90909 0.592  6  4   1   baseball
## 977       Jerry Sloan 24.68293 0.6030000 12.62500 0.485  2  0   1 basketball
## 236         Bobby Cox 27.82716 0.5560000 12.36364 0.493  5  1   1   baseball
```




```
## Top 10 coaches by Playoff Win Percentage
##                      N        GR    WP       PGR   PWP CC C HOF        S
## 7          Adam Walsh  1.312500 0.750 0.3333333 1.000  1 1   0 football
## 769  George Stallings 11.185185 0.495 0.3636364 1.000  1 1   0 baseball
## 774     George Wilson 10.000000 0.447 0.6666667 1.000  1 1   0 football
## 818        Hank Bauer  7.030864 0.522 0.3636364 1.000  1 1   0 baseball
## 1314       Lou Rymkus  1.187500 0.611 0.3333333 1.000  1 1   0 football
## 1574      Potsy Clark  7.375000 0.604 0.3333333 1.000  1 1   0 football
## 1879   Vince Lombardi  8.500000 0.738 3.3333333 0.900  5 5   1 football
## 780        Gil Hodges  8.728395 0.467 0.7272727 0.875  1 1   0 baseball
## 110     Bill Carrigan  6.191358 0.494 0.9090909 0.800  2 2   0 baseball
## 559    Don McCafferty  2.937500 0.620 1.6666667 0.800  1 1   0 football
```




```
## Top 10 coaches by Conference Championships
##                   N       GR        WP       PGR   PWP CC  C HOF          S
## 1559   Phil Jackson 20.00000 0.7040000 20.812500 0.688 13 11   1 basketball
## 328   Casey Stengel 23.24691 0.5080000  5.727273 0.587 10  7   1   baseball
## 1145    John McGraw 29.43827 0.5860000  4.909091 0.481 10  3   1   baseball
## 102  Bill Belichick 25.00000 0.6830000 14.333333 0.721  9  6   0   football
## 403     Connie Mack 47.87037 0.4860000  3.909091 0.558  9  5   1   baseball
## 1085   Joe McCarthy 21.52469 0.6150000  3.909091 0.698  9  7   1   baseball
## 1514      Pat Riley 23.21951 0.6360000 17.625000 0.606  9  5   1 basketball
## 1608   Red Auerbach 17.28049 0.6620000 10.500000 0.589  9  9   1 basketball
## 1722  Scotty Bowman 26.10976 0.6846450 22.062500 0.632  9  9   1     hockey
## 1807      Toe Blake 11.14634 0.6622517  7.437500 0.689  8  8   1     hockey
```




```
## Top 10 coaches by Championships
##                    N       GR        WP       PGR   PWP CC  C HOF          S
## 1559   Phil Jackson 20.00000 0.7040000 20.812500 0.688 13 11   1 basketball
## 1608   Red Auerbach 17.28049 0.6620000 10.500000 0.589  9  9   1 basketball
## 1722  Scotty Bowman 26.10976 0.6846450 22.062500 0.632  9  9   1     hockey
## 1807      Toe Blake 11.14634 0.6622517  7.437500 0.689  8  8   1     hockey
## 328   Casey Stengel 23.24691 0.5080000  5.727273 0.587 10  7   1   baseball
## 1085   Joe McCarthy 21.52469 0.6150000  3.909091 0.698  9  7   1   baseball
## 1521     Paul Brown 20.37500 0.6720000  5.666667 0.529  7  7   1   football
## 102  Bill Belichick 25.00000 0.6830000 14.333333 0.721  9  6   0   football
## 421   Curly Lambeau 23.75000 0.6310000  1.666667 0.600  6  6   1   football
## 754    George Halas 31.06250 0.6820000  3.000000 0.667  6  6   1   football
```

Some potential problems

- Dataset does not capture coach's personalities, effect on players, effect on the league, work in community, ...
- Many coaches are in HOF despite having less than a full year of head coaching.This is probably due to their assistant coaching experience which is not captured in this dataset
  - It would be nice to filter this dataset to coaches who have only over a year experience to account for outlier such as coaches with a win-loss percentage of 1 because they have only coached one game. However, this would remove almost 50 coaches who made the hall f fame despite having less than a year of head coaching experience

```
head(df[df$GR<1 & df$HOF==1,])
```

```
##                   N        GR        WP    PGR PWP CC C HOF        S
## 27      Al Spalding 0.7777778 0.6240000 0.0000 0.0  1 0   1 baseball
## 44        Alf Smith 0.2195122 0.6666667 0.3125 0.2  0 0   1   hockey
## 79  Barney Stanley 0.2804878 0.1904762 0.0000 0.0  0 0   1   hockey
## 88  Benny Friedman 0.8750000 0.3570000 0.0000 0.0  0 0   1 football
## 94   Bert Olmstead 0.7804878 0.2291667 0.0000 0.0  0 0   1   hockey
## 117    Bill Dickey 0.6481481 0.5430000 0.0000 0.0  0 0   1 baseball
```